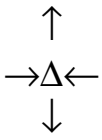
SIMDISS

Computer program - Computation of resemblance matrices and diversity indices

User's Manual

Version 2.0e (December, 1998)

Nico Salmaso



http://www.limno.eu/SimDiss

Padova, March 2001

Salmaso, N., 2001. SIMDISS. Computer program - Computation of resemblance matrices and diversity indices. User's Manual, V. 2.0e. http://www.limno.eu/SimDiss. 22 pp.

This program is freeware and is provided "as it stands", without warranty of any kind. No assurance as to accuracy, completeness or obtainable results are given, therefore the author is not obliged to provide the users with any assistance. The user assumes all risk for any damages arising in connection with the use and quality of this software.

STATISTICA is a trademark of StatSoft, Inc., 2300 East 14th Street, Tulsa, OK 74104, WEB: http://www.statsoft.com
SYSTAT is a trademark of SPSS Inc., WEB: http://www.spssscience.com/systat/
MICROSOFT EXCEL is a trademark of Microsoft Corp., WEB: http://www.microsoft.com
MICROSOFT WORD is a trademark of Microsoft Corp., WEB: http://www.microsoft.com
TURBO PASCAL is a trademark of Borland Software Corp., WEB: http://www.inprise.com

Nico Salmaso, PhD
IASMA Research and Innovation Centre
Istituto Agrario di S. Michele all'Adige
Via E. Mach, 1
I-38010, San Michele all'Adige, Trento (Italy)

E-mail: nico.salmaso@iasma.it

WEB
http://www.iasma.it
http://www.limno.eu

1. OVERVIEW

2. RESEMBLANCE COEFFICIENTS
- **2.1. Rationale**
- **2.2. Coefficients**
  - *2.2.1. Binary coefficients*
  - *2.2.2. Quantitative coefficients*
- **2.3. Relationships between pairs of coefficients**

3. COMPUTATION OF THE RESEMBLANCE COEFFICIENTS
- **3.1. Input operations**
- **3.2. Preliminary data transformations**
- **3.3. Computation of the resemblance matrices**
  - *3.3.1. Input/Output control (Option P)*
  - *3.3.2. Matrix transformations (Option M)*

4. OTHER PROCEDURES
- *4.1. Diversity indices*

Appendix 1 - Resemblance coefficients

Appendix 2 - Diversity coefficients

Appendix 3 - Files enclosed in SD_200e.EXE

References

1. OVERVIEW

SIMDISS is a computer program written in Turbo Pascal 7.01 for the computation of resemblance matrices. The objects to be compared may represent samples (quadrats), individual species or other different entities. The program may be used in different fields, but originally it has been written for the study of the temporal and spatial variations of biological communities. In addition, *SIMDISS* computes also various diversity indices.

In community ecology the data matrices are usually represented by tables showing the amount of different species in several samples. The majority of ecologists use input matrices whereby each row represents a single species and each column a single sampling unit (quadrat). In the example reported below a table with *s* species and *n* quadrats has been reported.

	Sample 1	Sample 2	. . .	Sample *n*
Species 1				
Species 2				
. . .				
Species *s*				

The practical illustration of the general characteristics of the program and the description of the different resemblance coefficients will refer to typical quantitative rectangular matrices where rows and columns represent species and samples, respectively.

The main menu of the program is reported in Fig. 1.

```
                          SIMDISS 2.0e
 ____________________________________________________________________

RESEMBLANCE COEFFICIENTS      ['Beta diversity']

B. Binary coefficients
T. Quantitative coefficients 1 ('Association coefficients')
D. Quantitative coefficients 2 ('Distance coefficents')
 ____________________________________________________________________

OTHER PROCEDURES

V. Diversity indices          ['Alpha diversity']
 ____________________________________________________________________


C. List of coefficients
F. Input/Output operations
I. About the program
Q. Quit

?:

Input matrix: ?   0X0
```

Fig. 1

2. RESEMBLANCE COEFFICIENTS

2.1. Rationale

A set of data, with *s* species and *n* samples, may be represented in the following matrix form:

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & X_{13} & \dots & X_{1j} & \dots & X_{1n} \\ X_{21} & X_{22} & X_{23} & \dots & X_{2j} & \dots & X_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ X_{i1} & X_{i2} & X_{i3} & \dots & X_{ij} & \dots & X_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ X_{s1} & X_{s2} & X_{s3} & \dots & X_{sj} & \dots & X_{sn} \end{bmatrix}$$

The basic input matrix **X** represents a rectangular array of numerical entries denoted by X_{ij}, where *i* refers to attributes (species, rows $i = 1,2,...,s$) and *j* refers to objects (samples, columns $j = 1,2,...,n$). Each row in **X** describes the distribution of the *i*th species along the considered samples, whereas each column reports the quantity of the different species in the *j*th sample. A row of **X** may be referred to as a *species vector* (or *row vector*) and a column as a *quadrat vector* (or *column vector*) (ORLOCI, 1978).

In community ecology each entry in **X** represents the result of an observation (species counting or other direct or derived quantities, e.g. weight, volumes, proportions, presence/absence etc.) of species *i* in samples *j*.

SIMDISS allows, starting from a rectangular matrix **X** of order $s \times n$, to compute many indices expressing the resemblance between every possible couple of column vectors. In this context the term *resemblance* is used to indicate the whole body of similarity/dissimilarity and distance indices (cf. ORLOCI, 1978). Likewise, SNEATH & SOKAL (1973: 116-120) use the term *similarity* to indicate the similarity measures in the strict sense of the word and the dissimilarity measures (distances may be considered measures of dissimilarity).

The computation of an index is carried out between all the possible pairs of samples (designated by *j* and *k*), for a total of $n(n-1)/2$ comparisons (excluding the self-comparisons). The results are saved in a symmetrical resemblance matrix whose elements S_{jk} represent the value of a particular index computed between the pairs of samples *j* and *k*. For example, if we consider the particular case of 5 samples and *s* species,

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & X_{13} & X_{14} & X_{15} \\ X_{21} & X_{22} & X_{23} & X_{24} & X_{25} \\ \dots & \dots & \dots & \dots & \dots \\ X_{s1} & X_{s2} & X_{s3} & X_{s4} & X_{s5} \end{bmatrix}$$

the resemblance matrix is represented as:

$$\mathbf{S} = \begin{bmatrix} S_{11} & S_{12} & S_{13} & S_{14} & S_{15} \\ S_{21} & S_{22} & S_{23} & S_{24} & S_{25} \\ S_{31} & S_{32} & S_{33} & S_{34} & S_{35} \\ S_{41} & S_{42} & S_{43} & S_{44} & S_{45} \\ S_{51} & S_{52} & S_{53} & S_{54} & S_{55} \end{bmatrix}$$

The resemblance matrices computed by *SIMDISS* may be used for further elaborations, for example as input matrices in commercial statistical computer programs for ordination purposes (e.g. multidimensional scaling) and classification (cluster analyses).

The progressive development of quantitative ecology – and other scientific disciplines (e.g. taxonomy, anthropology, psychology, etc.) which require the use of appropriate resemblance functions – has resulted in the rapid elaboration of various coefficients. The numerous resemblance functions have not found an adequate classification in single studies ("...any attempt at an exhaustive catalog of them would require many pages...", SNEATH & SOKAL, 1972: 129). In community ecology the classification and description of various coefficients has been carried out underlying, for example, their metric properties (ORLOCI, 1978) or their use in relation to the analysis Q (the association of pairs of samples on the basis of all species, i.e. the case considered in the above examples) or R (the association of pairs of species on the basis of all samples) (LEGENDRE & LEGENDRE, 1984).

For practical reasons, the resemblance coefficients implemented in *SIMDISS* have been subdivided into binary and quantitative coefficients (see Fig. 1). The first group of coefficients is used for the analysis of binary matrices (whose entries are represented by 0 and 1 to indicate species absence and presence, respectively), whereas the second group is used for the analysis of quantitative matrices. However, it is necessary to underline that the correct choice of a coefficient for the analysis of a particular dataset should be always motivated, in relation to further data elaboration or to the objectives of the analyses, e.g. Q or R type (see SALMASO, 1996). General criteria for the choice of appropriate coefficients in community ecology may be found in LEGENDRE & LEGENDRE (1984), ORLOCI (1978), FAITH *et al.* (1987). In particular, the problem of the inclusion of double zeroes in comparisons is widely discussed in LEGENDRE & LEGENDRE (1984); the distinction between metric and non-metric properties of different coefficients is discussed in ORLOCI (1978), PIELOU (1984) and LEGENDRE & LEGENDRE (1984).

2.2. Coefficients

On the whole, both binary and quantitative coefficients computed by the computer program may be classified in two broad groups: similarity coefficients and

dissimilarity/distance coefficients. Similarity coefficients have their maximum values when two samples are identical and the minimum values when two samples have no species in common. Similarity values may be transformed into distances, taking – as for coefficients ranging between 0 and 1 – their complement to one. The complete list, mathematical formulations and relationships among the resemblance coefficients implemented by *SIMDISS* are reported in Appendix 1 (see SNEATH & SOKAL (1972), ORLOCI (1978) and LEGENDRE & LEGENDRE, 1982 for details).

2.2.1. Binary coefficients

Resemblance values are determined by considering the number of common species in comparison to the number of exclusive species present in two of the samples being considered.

SIMDISS reports some of the most widely used similarity coefficients operating on binary matrices as well as their respective distances (dissimilarities). These latter values may be obtained using appropriate formulae (cf Appendix 1) or computing the complement to 1 of the single values.

For example, the similarity coefficient of *Jaccard* is computed using the following formula:

,

where, for every pair of column vectors, *a* is the number of common species, whereas *b* and *c* are the species present in the first and second sample, respectively. The complement to 1 of this index is the distance of *Marczewski-Steinhaus*:

$$\mathrm{DMS} = 1 - \mathrm{SJA} = 1 - \frac{a}{a+b+c} = \frac{b+c}{a+b+c}.$$

Analogous relationships tie the *Sorensen*'s similarity index (*Dice*) with the *Non-metric coefficient* and the *Sokal & Sneath*'s similarity with its complement to one.

2.2.2. Quantitative coefficients

The quantitative coefficients have been subdivided into two groups (Quantitative coefficients 1 and 2). The first group includes various similarity, dissimilarity and distance measures, whereas the second group comprises the set of distance measures related to the metric of *Minkowski* (including the *Euclidean* distance and the *absolute* or *City-block* distance).

In the first group the coefficients are reported as similarities, distances or both. In this latter case, the distances are computed considering the complement to 1 of the similarity values (or computing appropriate original formulae, cf Appendix 1). This is the case of the coefficients of *Ružička*, *Steinhaus*, *Gower* and *similarity "chi-square"*. As for the complement to 1 of the similarity of *Steinhaus*, the coefficient obtained is commonly identified with the name of *Bray & Curtis* index (BRAY & CURTIS, 1957).

However, the original formulation of this coefficient is anterior, having been reported as *percentage difference* by ODUM (1950) (LEGENDRE & LEGENDRE, 1982); this dissimilarity measure is reported also as *percentage dissimilarity* or *percentage distance* (GAUCH, 1982).

In some cases the coefficients considered in the first group have a direct correspondence with the binary coefficients. For example, when the data are binary, the similarities of *Ružička* and *Steinhaus* are equivalent to the similarities of *Jaccard* and *Sorensen*, respectively; likewise, the complement to 1 of the similarity of *Ružička* and the *Bray & Curtis* index have an identical correspondence with the distances of *Marczewski–Steinhaus* and the *non-metric coefficient*, respectively.

2.3. Relationships between pairs of coefficients

Many coefficients are characterised by more or less direct relationships. This became evident considering their direct comparison and the comparability of the results obtained by the subsequent processing of the respective resemblance matrices **S** (e.g. clustering and ordination). Essentially, three cases may be distinguished:

1. coefficients reported in scientific literature under different names and with different formulations, but giving equivalent results; this is the case of the Stander's similarity index (*SIMI*, in JOHNSON & MILLIE, 1982) which is equivalent to the *cosine separation* (in ORLOCI, 1978: 199);
2. monotonic coefficients; this type of relationship exists, e.g., between the coefficients of *Jaccard* and *Sorensen* and the coefficients of *Ružička* and *Steinhaus*;
3. coefficients characterised by a high degree of correlation, e.g. between the indices of *Bray & Curtis* and *Whittaker*.

Fig. 2 reports a matrix scatterplot illustrating the relationships between many of the distances implemented by *SIMDISS*. Each coefficient is represented by all the possible comparisons *n(n-1)/2* carried out on the basis of the phytoplankton density values (cells ml^{-1}) determined in 15 samples collected, during an annual cycle, in a small quarry lake (Appendix 3). Rare species, found on one occasion only, were not considered in the calculation. Moreover, logarithmic transformation ($Y_{ij}=\ln(X_{ij}+1)$) of the original data was applied to reduce the weight of the most abundant species. Computations have been carried out on an input matrix of 15 columns (samples) and 39 rows (phytoplanktonic *taxa*) giving a total of 105 comparisons for every coefficient.

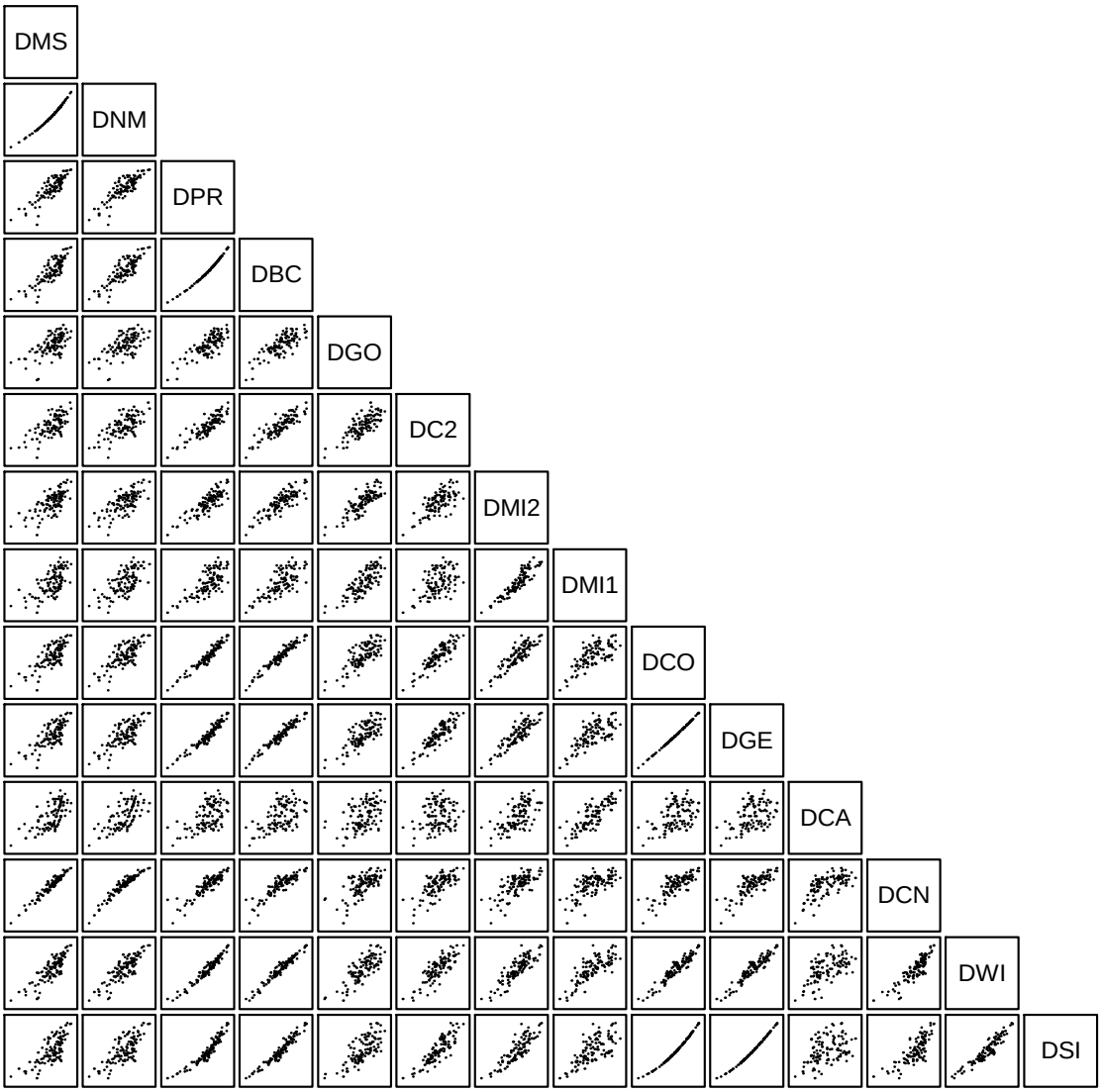


Fig. 2 – Relationship between some distance coefficients implemented by *SIMDISS*. DMS: *Markzewski-Steinhaus*; DNM: *non-metric coefficient*; DPR: *1-Ružička*; DBC: *Bray-Curtis (percentage difference)*; DGO: *1-Gower*; DC2: *"chi square" metric*; DMI2: *euclidean distance*; DMI1:*city-block distance*; DCO: *chord*; DGE: *geodesic*; DCA: *Canberra*; DCN: *Canberra, normalised*; DWI: *Whittaker*; DSI: *1-SIMI*.

3. COMPUTATION OF THE RESEMBLANCE COEFFICIENTS

3.1. Input operations

Input files must be in ASCII (text files). Columns and rows should contain the objects (e.g. samples) and descriptors (e.g. species); every column must report a heading (max 8 characters) for its univocal identification (e.g. sampling station and date). Column widths must be of 13 characters.

An example of matrix utilizable by *SIMDISS* is reported in the file PHYTDENS.PRN, which is included in the self extracting archive SD_200.EXE. The archive has been sorted by the number of species present in the different rows; it has a total of 15 columns (which represent the analysed samples) and 71 rows (species). A subset of this file, including only the species present at least in two samples (giving a total of 15 columns and 39 rows), has been utilised in the study of relationships among distances (Fig. 2). *SIMDISS* is able to read only subsets of rows (not columns); for example, with the file PHYTDENS.PRN it is possible to conduct computations on matrices with a number of columns and rows (col.×rows) of 15×71 (the whole dataset), 15×39, 15×10 and so on. At present the program can accept matrices with a maximum of 100 columns and 100 rows.

> ASCII files may be easily obtained from spreadsheet archives. An example is given with the file PHYTDENS.XLS (Excel 97-2000 and 5.0/95), enclosed in the self extracting archive, along with the computer program. Each column has a width of 13 characters. Starting from this archive, in Excel, an input file utilisable by *SIMDISS* may be obtained with the options *Save as* (from the menu *File*); in *File type* choose *Formatted text* (delimited by space, .PRN type). The enclosed ASCII file PHYTDENS.PRN has been obtained in this way.

Afterwards the options available in *SIMDISS* will be identified in courier, underlined.

The input files may be opened by choosing option F from the main menu. First of all, indication as to the path where to read a file from and to direct output to a specific directory is required. The directory must be inserted indicating the drive (hard disk or removable units) and the backlash, e.g.:

C:\SIMDISS\

Press enter. At this point it is necessary to indicate the input file, including the possible file extension, e.g.:

PHYTDENS.PRN

In the next step the number of columns and rows (headings not included) must be indicated. For example, if you want to open the whole set of data saved in PHYTDENS.PRN, you should specify 15 columns and 71 rows, whereas if you want to exclude the rare species, found only in one sample (cf. 2.2.3), you should specify 15 columns and 39 rows.

At this stage, the last information required is the name of the output file. You may insert a name (max 8 characters, with or without extension), or press on the empty field: in this case the name may be indicated later on (the temporary default name is OUT.PRN).

3.2. Preliminary data transformations

The following options allow data transformations, including normalisations and scale change, to be made.

In CONVERSION FACTOR each entry is multiplied by the number inserted in this field. Insert 1 if you do not want to change the original data matrix.

With the successive option (TRANSFORMATIONS/NORMALISATIONS), it is possible to transform the dataset. The available options include binary transformation and normalisations:

```
O. no transformations
B. binary 0-1
N. logn(Xi+1)              (natural logarithm)
D. log10(Xi+1)             (logarithm, base 10)
I. log2(Xi+1)              (logarithm, base 2)
2. square root
3. cube root
4. double square root
S. arcsin(sqrt(Xij))       (arcsin(√Xij), only for proportions, 0-1 range)
```

After these options, the program begins to read the data set, going back to the main menu; now, in the last row you should see the path and the name and size (columns × rows) of the input matrix.

N.B. From the principal menu it is possible to immediately choose a group of resemblance measures (options B, T, D) bypassing the preliminary reading of the data matrix with option F; in this case data loading will be recalled automatically.

3.3. Computation of the resemblance matrices

From the principal menu it is possible to choose three types of resemblance coefficients, subdivided into binary (option B) and quantitative (options T and D) coefficients; as for the computation of binary indices, the program automatically converts quantitative data into binary data. Every coefficient may be selected by a three letter code (both lower case and capital characters are accepted); the code begins with “S” and “D”, to indicate similarities and dissimilarities/distances, respectively.

For every distance, the metric properties (cf ORLOCI, 1978) are listed with option C (List of coefficients) in the principal menu.

At this point it is possible to compute the resemblance coefficients or to modify some parameters controlling the input/output operations and data matrix transformations (options P and N, sections 3.3.1. and 3.3.2., respectively).

3.3.1. Input/Output control (Option P)

This option allows modification of the various I/O operations:

```
O. Change the output file name
I. Load a new input file
D. Change the directory path

T. Matrix output type
E. Rectangular matrix format (Export to...)

S. Output on screen

U. Exit
```

O allows modification of the output file name. Starting from the same input data matrix, it is possible to compute different output resemblance data matrices;

I loads a different input data matrix;

D changes the directory path;

T controls the format of the output matrices; the following options are available:

```
T. Complete comparisons -> rectangular matrix
C. Complete comparisons -> vector column
S. Comparison between pairs of contiguous samples
   'beta turnover' -> AB, BC, CD...
P. Comparison between pairs of samples
   -> AB, CD, EF...
```

T: Symmetrical matrix (default); this type of matrix is used for successive data elaboration (ordination, cluster analysis).

C: Column. The whole set (*n(n-1)/2*) of comparisons between all the possible pairs of samples is saved in one single column. This allows easy comparison of different resemblance coefficients computed on the same input matrix.

S: Column. This option allows the comparison of pairs of contiguous samples. Let us suppose that a matrix with 4 samples identified with the headings A, B, C, D has to be analysed. *SIMDISS* computes the coefficients for the following comparisons: AB, BC and CD. This option is useful for the computation of the community change rate over temporal and spatial gradients (e.g. Salmaso, 1996). Even for this particular topic many interesting relationships among coefficients reported in literature need to be investigated. For example, the β-turnover reported in WILSON & SHMIDA (1984: 1057, β_T) is equivalent to the community turnover computed with the non metric coefficent, DNM.

P: Column. Comparison between pairs of samples. Taking into account the preceding example, *SIMDISS* computes the coefficients for the following comparisons: AB and CD.

E The symmetrical resemblance matrices may be saved (in ASCII format) in two different modalities, compatible with the format required by two commercial statistical packages: STATISTICA (STATSOFT, INC., 1997) (default) and SYSTAT (WILKINSON, 1990). In both cases it is necessary to convert the ASCII structure of the resemblance matrices in the format of the statistical packages.

> In the first case it is possible, for example, to easily import the resemblance matrix in *EXCEL*, and save the file in *.XLS* format (choose, from the menu, version 4.0); this file may be imported successively in STATISTICA. With recent versions of the program (e.g. 5.1), it is sufficient to access, sequentially, to the options *File*, *Import Data*, *Quick*, indicating the name of the file to import; at this point a new menu will appear "Quick import from Excel - Options". Before the confirmation of the import procedure be sure that the two options reported below ("Get case names..." and "Get variable names...") are checked.
>
> As for SYSTAT 5.0, DOS, the resemblance matrix may be imported directly as ASCII file, specifying, subsequently, in the EDIT module, the type of measure utilised in the computations. For example, as for a similarity matrix the appropriate command is TYPE=SIMILARITY; it is necessary to save the file (SAVE command) before leaving the EDIT module.
>
> **As for other versions of these programs, please refer to the respective user's manuals.**

S: permits the results of the comparisons to be shown on the computer screen (default).

3.3.2.Matrix transformations (Option M)

With option M it is possible to carry out the following operations:

```
Q. tranformation of column vectors (by column sum)
N. tranformation of column vectors (by column max)
M. tranformation of row vectors (by row max)
U. Exit
```

Q divides each element X_{ij} of a column vector by the sum of all its elements.

N divides each element X_{ij} of a column vector by the maximum value of all its elements.

M divides each element X_{ij} of a row vector by the maximum value of all its elements.

4. OTHER PROCEDURES

4.1. Diversity indices

The diversity indices computed by the program are:

Species_richness	Number of species (X_{ij}<>0)
Margalef	Margalef's index
Menhinik	Menhinik's index
Shannon_div	Shannon index (natural logarithm)
Shannon_eve	Shannon evenness
Simpson_pi	Simpson's index
Simpson_pi1	Simpson's index ("*finite community*")
Mc_Intosh	McIntosh's index
Mc_IntoshD	McIntosh "normalised" (independent from sample dimension)
Berger&Parker	Berger-Parker index

For a description of these indices and details about their computations see MAGURRAN (1988). The formulae implemented by *SIMDISS* are reported in Appendix 2. As for the input operations and the structure of the input files, details are described in the previous sections (3.1. and 3.2.). Diversity values are computed for every column *j*.

Some criteria for the correct choice and interpretation of diversity indices are reported in MAGURRAN (1988). However, diversity indices may be associated by strong relationships. A direct comparison between the pairs of coefficients computed by *SIMDISS* is reported in Fig. 3. Computations have been carried out utilising, in input, the matrix PHYTDENS.PRN (15 rows ×71 columns, original data, without preliminary transformations; see Appendix 3 for the description of the data matrix).

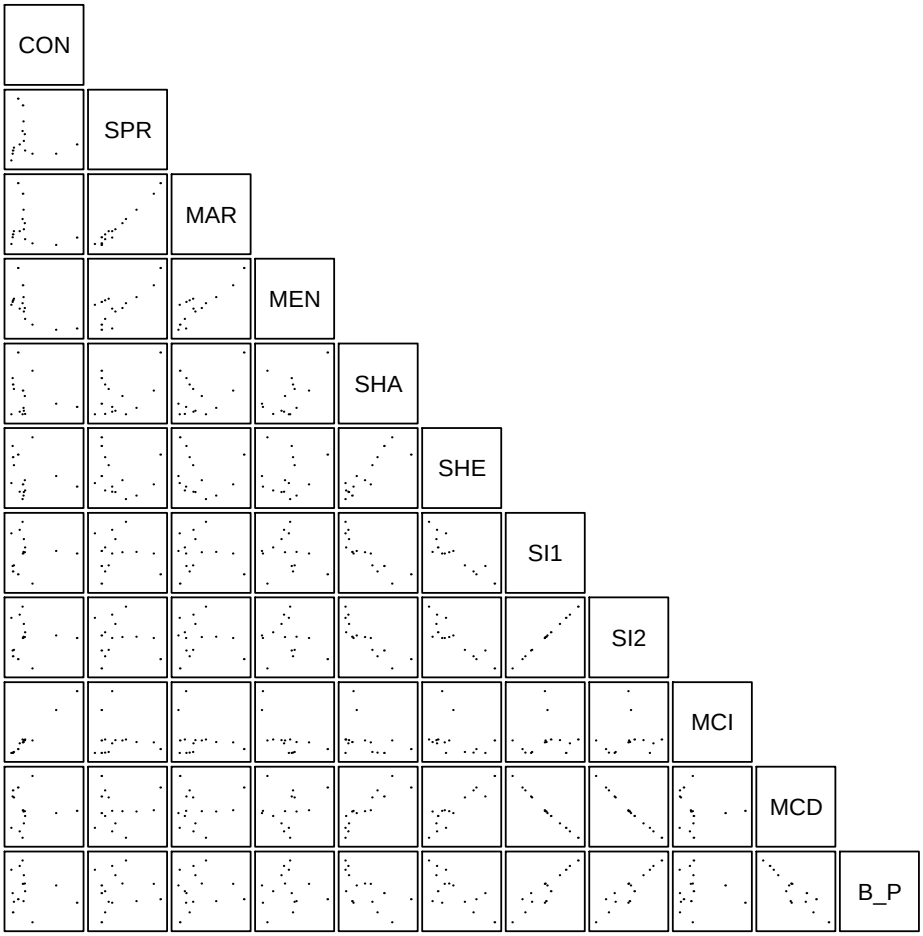


Fig. 3 - Relationship between the diversity coefficients (plus total density) implemented by *SIMDISS*. CON: *total density*; SPR: *Species richness*; MAR: *Margalef*; MEN: *Menhinik*; SHA: *Shannon diversity*; SHE: *Shannon evenness*; SI1: *Simpson p*; SI2: *Simpson p1*; MCI: *McIntosh*; MCD: *McIntosh D*; B_P: *Berger & Parker*.

Appendix 1 - Resemblance coefficients

a) Binary coefficients

Explanation of symbols.
a: number of common species in the two samples;
b and *c*: number of species present exclusively in the first and second sample, respectively.

- *Jaccard:* $\mathrm{SJA} = \frac{a}{a+b+c}$

- *Sorensen (Dice)* $\mathrm{SSO} = \frac{2a}{2a+b+c}$

- *Sokal & Sneath:* $\mathrm{SSS} = \frac{a}{a+2b+2c}$

- *Marczewski-Steinhaus:* $\mathrm{DMS} = 1 - \mathrm{SJA} = 1 - \frac{a}{a+b+c} = \frac{b+c}{a+b+c}$

- *Non metric coefficient*: $\mathrm{DNM} = 1 - \mathrm{SSO} = 1 - \frac{2a}{2a+b+c} = \frac{b+c}{2a+b+c}$

- *1-SSS*: $\mathrm{DSS} = 1 - \mathrm{SSS} = 1 - \frac{a}{a+2b+2c} = \frac{2b+2c}{a+2b+2c}$

b) Quantitative coefficients (1)

Explanation of symbols.
s: total number of species (variables);
i: species (row) index, 1...*s*;
j, *k*: sample (column) indices;

$$p_{ij} = X_{ij} \Big/ \sum_{i=1}^{s} X_{ij} \; ; \qquad p_{ik} = X_{ik} \Big/ \sum_{i=1}^{s} X_{ik}$$

- *Ružička*: $\mathrm{SRU} = \frac{\sum_{i=1}^{s} \min(X_{ij}, X_{ik})}{\sum_{i=1}^{s} \max(X_{ij}, X_{ik})}$

- *Steinhaus*: $$\mathrm{SST} = 2\frac{\sum_{i=1}^{s}\min(X_{ij}, X_{ik})}{\sum_{i=1}^{s}(X_{ij} + X_{ik})}$$

- *Gower*: $$\mathrm{SGO} = \frac{\left(\sum_{i=1}^{s} w_{ijk} \cdot S_{ijk}\right)}{\sum_{i=1}^{s} w_{ijk}};$$

for quantitative matrices: $S_{ijk} = 1 - \left(\left|X_{ij} - X_{ik}\right| / R_i\right)$ and $w_{ijk}=1$ (no double zeroes). R_i is the range of variation for species *S*.

- *Morisita*: $$\mathrm{SMO} = \frac{2 \cdot \sum_{i=1}^{s} X_{ij} \cdot X_{ik}}{\left(\lambda_j + \lambda_k\right) \cdot \sum_{i=1}^{s} X_{ij} \cdot \sum_{i=1}^{s} X_{ik}}$$, where:

$$\lambda_j = \sum_{i=1}^{s} X_{ij} \cdot \left(X_{ij} - 1\right) \Big/ \sum_{i=1}^{s} X_{ij} \cdot \left(\sum_{i=1}^{s} X_{ij} - 1\right), \text{ and}$$

$$\lambda_k = \sum_{i=1}^{s} X_{ik} \cdot \left(X_{ik} - 1\right) \Big/ \sum_{i=1}^{s} X_{ik} \cdot \left(\sum_{i=1}^{s} X_{ik} - 1\right)$$

- *"chi square similarity"*: $$\mathrm{SC2} = 1 - \mathrm{DC2} = 1 - \sqrt{\sum_{i=1}^{s} \frac{1}{S_i}\left(\frac{X_{ij}}{\sum_{i=1}^{s} X_{ij}} - \frac{X_{ik}}{\sum_{i=1}^{s} X_{ik}}\right)^2},$$

where S_i is the sum of the *i*th row (species) (LEGENDRE & LEGENDRE, 1984).

- 1- SRU (LEWANDOWSKY (1972)): $$\mathrm{DPR} = 1 - \mathrm{SRU} = 1 - \frac{\sum_{i=1}^{s}\min(X_{ij}, X_{ik})}{\sum_{i=1}^{s}\max(X_{ij}, X_{ik})}$$

- *Bray-Curtis*: $$\mathrm{DBC} = 1 - \mathrm{SST} = 1 - 2\frac{\sum_{i=1}^{s}\min(X_{ij}, X_{ik})}{\sum_{i=1}^{s}(X_{ij} + X_{ik})} = \frac{\sum_{i=1}^{s}\left|X_{ij} - X_{ik}\right|}{\sum_{i=1}^{s}(X_{ij} + X_{ik})}$$

- *1-Gower*: $$\mathrm{DGO} = 1 - \mathrm{SGO}$$

- *"chi square metric"*: $$DC2 = \sqrt{\sum_{i=1}^{s} \frac{1}{S_i} \left(\frac{X_{ij}}{\sum_{i=1}^{s} X_{ij}} - \frac{X_{ik}}{\sum_{i=1}^{s} X_{ik}} \right)^2}$$; cf SC2.

c) Quantitative coefficients (2)

Explanation of symbols.
See quant. coefficients 1

- *Minkowski metric*: $$DMI = \left(\sum_{i=1}^{s} \left| X_{ij} - X_{ik} \right|^r \right)^{1/r}$$;

  r=1: *city-block distance; r*=2: *euclidean distance*.

- *chord*: $$DCO = \sqrt{2 \left(1 - \frac{\sum_{i=1}^{s} X_{ij} \cdot X_{ik}}{\sqrt{\sum_{i=1}^{s} X_{ij}^2 \cdot \sum_{i=1}^{s} X_{ik}^2}} \right)}$$

- *geodesic*: $$DGE = \arccos \left(1 - \frac{DCO}{2} \right)$$

- *SIMI*: $$SSI = \frac{\sum_{i=1}^{s} p_{ij} \cdot p_{ik}}{\sqrt{\sum_{i=1}^{s} p_{ij}^2 \cdot \sum_{i=1}^{s} p_{ik}^2}}$$

- 1-*SIMI*: $$DSI = 1 - SSI$$

- *Canberra*: $$DCA = \sum_{i=1}^{s} \left(\frac{\left| X_{ij} - X_{ik} \right|}{X_{ij} + X_{ik}} \right)$$

- *Canberra, normalised*: $$DCN = \frac{1}{s^*} \cdot DCA$$

DCA must exclude double zeros in order to avoid indetermination. Therefore in the computation of this version of DCN, s^* is evaluated taking into consideration only the number of those variables (rows, species) having non-zero values in at least one of the couple of objects (columns, samples) under comparison. The same criteria were also used in previous versions of SIMDISS.

- *Whittaker*: $$\mathrm{DWI} = 0.5 \cdot \sum_{i=1}^{s} \left| \frac{X_{ij}}{\sum_{i=1}^{s} X_{ij}} - \frac{X_{ik}}{\sum_{i=1}^{s} X_{ik}} \right|$$

Appendix 2 - Diversity coefficients

Diversity indices (Magurran, 1988)

Explanation of symbols.
See quant. coefficients 1. Other symbols:

N, total density, $N = \sum_{i=1}^{s} X_{ij}$;

N_{max}: density of the most abundant species.

- *Total density (*CON*)* N
- *Species' richness (*SPR*)* s
- *Margalef's index (*MAR*)* $D_{Mg} = \frac{s-1}{\ln(N)}$
- *Menhinik's index (*MEN*)* $D_{Mn} = \frac{s}{\sqrt{N}}$
- *Shannon index (*SHA*)* $H' = -\sum_{i=1}^{s} p_{ij} \ln p_{ij}$
- *Shannon evenness (*SHE*)* $E = \frac{H'}{H_{max}} = \frac{H'}{\ln s}$
- *Simpson's index (*SI1, SI2*)* $D = \sum_{i=1}^{s} p_{hj}^{\ 2}$; (infinitely large community, SI1)

  $D = \sum_{i=1}^{s} \left(\frac{X_{ij}(X_{ij}-1)}{N(N-1)} \right)$; (finite community, SI2)
- *McIntosh's index (*MCI, MCD*)* $U = \sqrt{\sum_{i=1}^{s} X_{ij}^{\ 2}}$; (MCI)

  $D = \frac{N-U}{N-\sqrt{N}}$; (MCD)
- *Berger-Parker index(*B_P*)* $d = \frac{N_{max}}{N}$

Appendix 3 - **Files enclosed in SD_200e.EXE**

SIMDISS.EXE. SimDiss 2.0e(01), computer program file.

PARAMSD.MSD. Parameter file. This file saves the input/output configuration used during the last work session.

PHYTDENS.PRN. Example of input matrix file, 15 columns × 71 rows. The entries represent phytoplankton density values (cells ml^{-1}) determined in 15 water samples collected, during an annual cycle, in a littoral station of a small quarry lake in the province of Padova (NE Italy; see SALMASO *et al.*, 1995).

PHYTDENS.XLS. Example of input matrix file in Excel format (Excel 97-2000 and 5.0/95).

SDMANUAL.PDF. User's Manual.

README.TXT.

References

BRAY, J.R. & J.T. CURTIS, 1957. An ordination of the upland forest communities of Southern Wisconsin. *Ecol. Monogr*. 27: 325-349.

FAITH, D.P., P.R. MINCHIN & L. BELBIN, 1987. Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio* 69: 57-68.

GAUCH, H.G., 1982. *Multivariate analysis in community ecology*. Cambridge University Press, Cambridge.

JOHNSON, B.E. & D.F. MILLIE, 1982. The estimation and applicability of confidence intervals for Stander's Similarity Index (SIMI) in algal assemblage comparisons. *Hydrobiologia* 89: 3–8.

LEGENDRE, L. & P. LEGENDRE, 1984. *Écologie numérique. La structure de données écologiques, 2*. Masson, Paris-Presses de l'Université du Quebec.

LEWANDOWSKY, M., 1972. An ordination of phytoplankton populations in ponds of varying salinity and temperature. *Ecology* 53: 398-407.

MAGURRAN, A.E., 1988. *Ecological diversity and its measurement.* Croom Helm, London.

ODUM, E.P., 1950. Bird populations of the highlands (North Carolina) plateau in relation to plant succession and avian invasion. *Ecology* 31: 587–605.

ORLÓCI, L., 1978. *Multivariate analysis in vegetation research.* W. Junk B.V. Publishers, Boston.

PIELOU, E.C., 1984. *The Interpretation of Ecological Data: a Primer on Classification and Ordination.* John Wiley & Sons, New York.

SALMASO, N., 1996. Seasonal variation in the composition and rate of change of the phytoplankton community in a deep subalpine lake (Lake Garda, Northern It-aly). An application of nonmetric multidimensional scaling and cluster analysis. *Hydrobiologia* 337: 49-68.

SALMASO, N., M. MANFRIN & P. CORDELLA, 1995. Struttura e dinamica della comunità fitoplanctonica in un piccolo lago di falda (Rubàno, Padova). *S.It.E. Atti* 16: 703-706.

SNEATH, P.H.A. & R.R. SOKAL, 1973. *Numerical Taxonomy*. Freeman, San Francisco.

StatSoft, Inc., 1997. *STATISTICA for Windows* [Computer program manual]. Tulsa, OK: StatSoft, Inc., 2300 East 14th Street, Tulsa, OK 74104, phone: (918) 749-1119, fax: (918) 749-2217, email: info@statsoft.com, WEB: http://www.statsoft.com

WILSON, M.V. & A. SHMIDA, 1984. Measuring beta diversity with presence-absence data. *J. Ecol.* 72: 1055-1064.

WILKINSON, L., 1990. *SYSTAT: The System for Statistics.* SYSTAT, Inc., Evanston.